European Research Studies
Volume V, Issue (3-4), 2002, pp.
13-23

Algorithms for Computing the Regression Coefficients: Recommendations

Vassilis C. Hombas*

Abstract

We present a survey of possible algorithms and their rounding off tranca-tion, arithmetic error bounds. Experimental results confirm these errors and illustrate the dangers of some algorithms because of errors in the means. Specific recommendations are made as to which algorithms should be used.

1. Introduction

The problem of computing the regression coefficients: the gradient b and the constant b_0 as well as the coefficient of determ-ination R^2 and the elasticity of Y in relation to X: $E(Y,X)=b\overline{x}/\overline{y}$, reduces to computing the covariance and variance of a sample of N data points X_i, $i=1,2,\ldots,N$. It is one that seems, at first glance to be almost trivial but in fact can be quite difficult, particularly when N is large, and the covariance-variance is small and when we have by necessity errors in computing the means.

Let $\overline{X},\overline{Y}$ be the true means and $\hat{X},\hat{Y}$ be the estimated means after some possible errors rounding off, truncation or arithmetic, then

$$\hat{X}=\overline{X}+e_1 \text{ and } \hat{Y}=\overline{Y}+e_2$$

where e_1,e_2 may assume any real values. In the computation of the estimated regression coefficients by O.L.S. method, we use either of the following formulae:

Two-Pass Algorithm 1 (Abbreviated TPA1)

* Associate Professor, University of Athens, Department of Economics.

$$\left.\begin{array}{ll} b = NS_{xy} / NS_x^2 \qquad , & b_0 = \overline{Y} - b\overline{X} \\ \text{where } NS_{xy} = \Sigma\left(X - \overline{X}\right)\left(Y - \overline{Y}\right) \text{ and } NS_x^2 = \Sigma\left(X - \overline{X}\right)^2 & \\ E\left(Y,X\right) = b \cdot \overline{X} / \overline{Y} \quad , & R^2 = \dfrac{Nb_0\overline{Y} + b\Sigma XY - N\left(\overline{Y}\right)^2}{\Sigma Y^2 - N\overline{Y}^2} \end{array}\right\} \qquad (1)$$

where $E\left(Y,X\right)$ and R^2 are the elasticity of Y in relation to X and the determination's coefficient respectively.

Two-Pass Algorithm 2 (Abbreviated TPA2)

$$\left.\begin{array}{ll} b = NS_{xy} / NS_x^2 & b_0 = \overline{Y} - b\overline{X} \\ \text{where } NS_{xy} = \Sigma XY - N\overline{X}\overline{Y} \text{ and } NS_x^2 = \Sigma X^2 - N\overline{X}^2 & \\ E\left(Y,X\right) = b \cdot \overline{X} / \overline{Y} \quad , & R^2 = \dfrac{Nb_0\overline{Y} + b\Sigma XY - N\left(\overline{Y}\right)^2}{\Sigma Y^2 - N\overline{Y}^2} \end{array}\right\} \qquad (2)$$

One-Pass Algorithm 3 (Abbreviated OPA3)

$$\left.\begin{array}{ll} b = N^2S_{xy} / N^2S_x^2 \qquad , & Nb_0 = \Sigma Y - b\Sigma X \\ \text{where } N^2S_{xy} = N\Sigma XY - \left(\Sigma X\right)\left(\Sigma Y\right) \text{ and } N^2S_x^2 = N\Sigma X^2 - \left(\Sigma X\right)^2 & \\ E\left(Y,X\right) = b\Sigma X / \Sigma Y \quad , & R^2 = \dfrac{Nb_0\Sigma Y + b \cdot N\Sigma XY - \left(\Sigma Y\right)^2}{N\Sigma Y^2 - \left(\Sigma Y\right)^2} \end{array}\right\} \qquad (3)$$

It has already been indicated by Lynch (1988) first and Hombas (1991) after that the use of formulae involving mean values can lead to incorrect statistical values. Yet, it is very easy to prove that formulae (2) and (3) derive algebraïcally from (1).

Formula (3) is simply a variant of (2). Also it has been proved by the author that errors in the estimation of means lead to the loss of the algebraic equivalence of (1) and (2), as well as to a disagreement in the accuracy of their results. The most important is that it has been shown that all of the above problems can be avoided by using formulae OPA3 not counting $\overline{X}$ and $\overline{Y}$ which as we mentioned above is a simple variant of (2). The formulae TPA1 and TPA2 required passing through the data twice: once to compute $\overline{X}$ and $\overline{Y}$ and then again to compute NS_{xy} and NS_x^2. This may be undesirable in many applications, for example when the data sample is too large to be stored in main memory or when the covariance-variance is to be calculated dynamically as the data are collected. To avoid the two-pass nature of (1) and (2), it is standard practice to manipulate the definition of NS_x^2 and NS_{xy} into the OPA3 form. This form is suggested in statistical computing. Unfortunately although (3) is mathematically equivalent to (1) and (2), numerically it can be disastrous when the quantities ΣX^2 and $\left(\frac{1}{N}\right)(\Sigma X)^2$ may be very large in practice and will generally be computed with some rounding errors. If the variance is small, these numbers should cancel out almost completely in the subtraction of (3). Many of the correctly computed digits will be canceled leaving a computed NS_x^2 or NS_{xy} with a possible unacceptable relative error. The computed NS_x^2 can even be negative, a blessing in disguise since this alerts at least the programmer that disastrous cancellation has occurred. Also R^2 can even be out of the range (0,1). To avoid these difficulties, several alternatives one-pass algorithms have been introduced. These include the *updating* algorithms of Youngs and Cramer (1971), West (1979) and the *pairwise algorithm* of Chan et all (1979).

Of course the computing of NS_{xy} or NS_x^2 on a computer with machine accuracy u% may have percentage errors as large as $hu\%$ regardless of that algorithm is used. The value hu can be used as a criterion by which the accuracy of various algorithms is critisized, especially since error margins are functions only of h, u and N can often be derived: see Chan et all (1979).

In contrast to TPA1 and TPA2, the OPA3 enables us to calculate the estimate regression coefficients as well as the elasticity and the multiple coefficient of determination R^2 without precalculating the means. Thus, despite the fact that the TPA2 and OPA3 are obviously algebraïcally equivalent and almost identical in form, we will see that they should yield different results, because of the absence of $\overline{X}, \overline{Y}$, in (3). This will be illustrated in the example below. Note that when a number in Table 1 and 3 has a* on it, means incorrect estimate.

2. Hypothetical Example

We calculate the regression coefficients, the elasticity $E(Y,X)$ and the coefficient of determination R^2 for the distribution given by

x :	10	20	30	40	50
y :	8	12	15	21	24

In Table 1, five calculations for a given set of data are presented. In the first case, both errors are zero and as it is shown all three formulae give the correct results, $b = 0.41$, $b_0 = 3.7$, $E(Y,X) = 0.76$ and $R^2 = 0.99$.

Therefore, $E(Y/X) = 3.7 + 0.41x$. In the second case, where the errors are $e_1 = 0$ and $e_2 = 1$, the regression coefficient $b = 0.41$ is calculated correctly both by the TPA1 and the OPA3, while the intercept constant b_0, the elasticity $E(Y,X)$ and the coefficient of determination R^2 are calculated incorrectly. It is remarkable that $b_0 = 9.2$ which comes from TPA2, is very different from the correct result $b_0 = 3.7$. The example shows that an error of almost 148% in the estimate b_0 can arise from a 6% error in one mean and zero in the other mean. The TPA1 gives about 27% error in b_0, a fact which is reflected in a shift of the regression line. Likewise the TPA2 gives $b = 0.26$, that is an error of almost 37% in the regression coefficient (gradient) b, which is reflected in a change of the direction of the regression line. Moreover, the TPA1 gives $E(Y,X) = 0.72$, while the TPA2 gives $E(Y,X) = 0.45$ further apart from the correct estimates $E(Y,X) = 0.76$. Furthermore, the TPA1 gives incorrect result in the elasticity, but it is more accurate than formula (2). The values $R^2 = 21.32$ and $R^2 = 13.52$ which came from the TPA1 and the TPA2 respectively are not only false, but are also out of the range of R^2 which is $0 \le R^2 \le 1$. The correct estimate $R^2 = 0.98$ comes only from the OPA3. All of the above confirms the fact that the TPA1 is more accurate than the TPA2.

In <u>the third case</u> the errors are $e_1 = 1$ and $e_2 = 0$, the estimated means are $\hat{X} = 31$ and $\hat{Y} = 16$. The TPA1 gives erroneous values in all of our statistical parameters except in the case $E(Y,X) = 0.761$, which by chance has an almost correct value at the first two digits. This example shows that, when we use the TPA1 an error of 5% and 0.1% in b and b_0 respectively emerges from a 3% error in the mean of the independent variable X and a 0% error in the mean of Y. The coefficient of determination has a 30% error. A first glance at the TPA2 shows an overestimation $b = 0.47$ and an underestimation $b_0 = 1.43$ in relation to the correct results, that is an error of almost 13% and 61% respectively. The elasticity $E(Y,X) = 0.91$ and the coefficient of determination have errors 16% and 8% respectively. Again we see that the accuracy from the TPA1 is better than the one coming from the TPA2.

In <u>the fourth case</u> the errors are $e_1 = -2$ and $e_2 = 1$ and the estimated means are $\hat{X} = 28$ and $\hat{Y} = 17$. The TPA1 gives incorrect results $b = 0.39$ and $b_0 = 6.02$, $E(Y,X) = 0.64$ and $R^2 = 32.5$. The example shows that an error of almost 5%, 38% and 18% for the first three statistics respectively (The estimate $R^2 = 32.5$ is out of its range) can arise from an error of almost 6% in both means. The TPA2 gives $b = 0.27$ and $b_0 = 9.44$, $E(Y,X) = 0.44$. This shows an error of about 52%, 61% and 72% respectively. Therefore, again, the example points out that the TPA2 is less accurate than that the TPA1. The OPA3 continues to give correct results.

Now let's come <u>to case 5</u>. Here we have increased the numerical values of the X variable by 2 so that the relation $e_2 : e_1 = \bar{Y} : \bar{X}$; $b_T = 0.41$ is satisfied. This case is admittedly difficult to happen but not impossible. The errors here are $e_1 = 2$ and $e_2 = 1$ and satisfy the relation $1 : 2 = e_2 : e_1 = \bar{Y} : \bar{X}$; 0.41. In fact, the TPA1 gives here "correct" results $b = 0.41, b_0 = 3, E(Y,X) = 0.82$. The slight disagreement is due to the fact that part of our hypothesis is not fully satisfied. The value $R^2 = 5.37$ lies out of its range because the formula contains $\bar{Y}$. The TPA2 gives $b = 0.24$, $b_0 = 8.84$ and

$E(Y,X)=0.48$ i.e. wrong results with enormous percentage errors 71%, 61% and 71% respectively, emerging from an error 6% in both means. Here, too, the TPA1 gives better results than (2). Formula (3) gives correct and exact results.

3. Real Example

Baczkowski (1989) had an experience in buying a caravan that inspired and led him to some statistical work. In an exhibition of cars of this sort he got an unexpectedly interesting set of data suitable for the use of the OLS method in regression. The data are shown below in Table 2.

Of interest is the question whether the resale price is influenced by any extra accessories such as a heater or a fridge, or it is solely determined by the age of the caravan. The data set leads to a study of the choice of regression model. After trying some models, Baczkowski realized that a linear approximation in the regression suffered from the presence of heteroscedasticity in the data, while a quadratic approximation had extrapolation problems. Finally taking logarithms of prices and regressing them on age he found (see Fig. 1)

$$\log p(x)=3.79-(0.0563)X \text{ , where } X=\text{age} \quad (4)$$

from which he obtained

$$p(x)=P_0(0.878)^x,\ x=0,1,2,3... \quad (5)$$

From (5) we realize easily that on average a caravan loses 12.2% or about 12% of its value each year. Inflation had not been included in any of these calculations. Now let' us come to our case. For convenience let's call $\log_{10}(\text{price})=Lp$. In this real example $\bar{X}=4.5$, $\bar{L}p=3.537384$. A truncation of five digits of the second mean gives $\tilde{L}p=3.5$, then we have an error of almost 1%. All the calculations are given in the following Table 3.

Table 3 shows calculations for a given set of the Baczkowski data. From the TPA1 the regression coefficient b is calculated correctly, whereas the intercept constant b_0 is computed incorrectly with an error of about 2% arising from an error of almost 1% in the mean of log (prices) and zero in the mean of age. The calculations using TPA2 are all incorrect. A percentage error of 14% in the direction of the regression line and about 2% in shifting arises from an 1% error in the mean of log (prices) only.

Using the wrong estimates we ended up with the wrong exponential model:

$$p = 5128(0.895)^x \tag{6}$$

compared with the correct one (Baczkowski)

$$p = 6166(0.878)^x \tag{7}$$

(see Fig. 2)

Comparing these two models we note that according to Baczkowski (7) the devaluation is on average 12,2% while in the wrong model it is almost 10,5% which is a considerable error. Further, for a certain age of a caravan, the longitude of its value is narrow. This is probably an indication that any additional accessories have little result in the resale price of the used caravan. Another explanation is that old caravans might have fewer accessories, even if for a certain age, the above accessories practically do not appear to cause an increase in the value of the caravan. For more information and conclusions of the basis of these data i.e. particular devaluations and whether the salesmen have the same depreciation curve, the reader should refer to the article by Baczkowski.

4. Concluding Remarks

The results of the previous sections provide a basis for making an intelligent choice of algorithm for accurately computing the regression coefficients b and b_0 and the other statistics which join them as the multiple determination's coefficient R^2 and standard errors $\hat{\sigma}_{b_0}$ and $\hat{\sigma}_b$ as well as the elasticity of Y with respect to X. In all situations the one-pass algorithm 3 can be recommended as it stands. If the data consist only of integers, small enough that no overflows occur, then OPA3 should be used with the sums computed in integer arithmetic. In this case no round off errors occur until the final step of combining the two sums, in which a division by N occurs.

Consequently, even if TPA1 is better than the TPA2 with respect of exactness and correctness, neither of those should be used towards finding the above statistical parameters in modern statistics since their use in combination with cheap calculations is restricted. Yet, in TPA2 and OPA3 the calculation of the rolling mean or current values ΣX of ΣY progresses with the entry of the data. The TPA2 often gives wrong results, while its variant to OPA3 always gives correct results. Therefore, exclusive of machine accuracy, the one-pass Algorithm 3 should be preferred and a Lemma tactic would be the following: *Whenever we cope with an algorithm involving mean values, transform the formula in such a way that these mean values are removed.*

References

Baczkowski, J. A., (1989), "On Buying a Caravan", *Teaching Statistics* 11(3), 66-68.

Chan, T. F., Golub, G. H. and Leveque, R. J. (1979), "Updating formulae and a Pairwise Algorithm for Computing Sample Variances", *Compast 1982,* Proceedings of the 5th Symbosiumheld at Toulouse, eds. H. Caussiuus et al., 30-41.

Hombas, C. V., (1991), "Accurate Computation of Covariance", *Teaching Statistics*, 13(3), 87-89.

Lynch, R. M., (1988), "The Variance with Computational Error in the Mean", *Teaching Statistics*, 10(2), 46-48.

West, P. H. P., (1998), "Updating Mean and Variance Estimates An Improved Method", *Communications of the Association for Computing Machinery*, 22, 532-535.

Younds, E. A. and Cramer, E. M., (1971), "Some Results Relevant to Choice of Sum and Sum-of-Product Algorithms", *Technometrics*, 13, 657-665.

Table 1: *Computation of the regression coefficients b, b_0, the elasticity E(Y,X) and the coefficient of determination R^2.*

Case 1	**Case 2**
$\bar{X}=30 \quad \bar{Y}=16$	$\bar{X}=30 \quad \bar{Y}=16$
$e_1=0 \quad e_2=0$	$e_1=0 \quad e_2=1$
$\hat{\bar{X}}=30 \quad \hat{\bar{Y}}=16$	$\hat{\bar{X}}=30 \quad \hat{\bar{Y}}=17$
	(percent. error $\cong$6% in $\bar{Y}$)
$N=5,\ \Sigma X=150,\ \Sigma Y=80,$	$N=5,\ \Sigma X=150,\ \Sigma Y=80,$
$\Sigma XY=2810,\ \Sigma X^2=5500$	$\Sigma XY=2810,\ \Sigma X^2=5500$
$NS_x^2=1000,\ NS_y^2=170,$	$NS_x^2=1000,\ NS_y^2=175,$
$NS_{xy}=410,\ \Sigma Y^2=1450$	$NS_{xy}=410,\ \Sigma Y^2=1450$

TPA1 $b = \frac{NS_{xy}}{NS_x^2} = \frac{410}{1000} = 0.41$, $b_0 = \bar{Y} - b\bar{X} = 3.7$ $E(y,x) = (0.41)30/16 = 0.76,$ $R^2 = 0.988$ all are correct estimates	**TPA1** $b = \frac{NS_{xy}}{NS_x^2} = \frac{82}{200} = 0.41$, $b_0 = \hat{Y} - b\hat{X} = 4.7^*$ (p.e 27%) $E(Y,X) = (0.41)30/17 = 0.72^*,$ $R^2 = 21.32^*$ only one correct estimate
TPA2 $b = \frac{S_{xy}}{S_x^2} = \frac{2810/5 - 30\cdot 16}{5500/5 - 30^2} = 0.41$, $b_0 = 3.7,\ E(Y,X) = 0.76, R^2 = 0.988$ all are correct estimates	**TPA2** $b = \frac{2810/5 - 30\cdot 17}{5500 - 30^2} = 0.26^*$, $b_0 = 17 - (0.26)309.2^*$ (p.e 148%) $E(Y,X) = 0.26\ 30/17 = 0.45^*, R^2 = 13.5^*$ all are incorrect estimates
OPA3 $b = \frac{25S_{xy}}{25S_x^2} = \frac{5\cdot 2810 - 150\cdot 80}{5000} = 0.41$ $5b_0 = \Sigma Y - b\Sigma X = 80 - (0.41)150 = 18.5$, $b_0 = 3.7,\ E(Y,X) = (0.41)150/80 = 0.76,$ $R^2 = 0.988$ all are correct estimates	**OPA3** $b = \frac{25S_{xy}}{25S_x^2} = \frac{5\cdot 2810 - 150\cdot 80}{5000} = 0.41$ $5b_0 = 80 - (0.41)150 = 18.5$, $b_0 = 3.7$ $E(Y,X) = (0.41)150/80 = 0.76,$ $R^2 = 0.988$ all are correct estimates

Table 1 (continued): *Computation of* $b, b_0, E(Y,X), R^2$

Case 3	**Case 4**
$\bar{X} = 30 \quad \bar{Y} = 16$ $e_1 = 1 \quad e_2 = 0$ (p.e = 3%) $\hat{X} = 31 \quad \hat{Y} = 16$	$\bar{X} = 30 \quad \bar{Y} = 16$ $e_1 = -2 \quad e_2 = 1$ (both p.e = 6%) $\hat{X} = 28 \quad \hat{Y} = 17$
$N = 5,\ \Sigma X = 150,\ \Sigma Y = 80,$ $\Sigma XY = 2810,\ \Sigma X^2 = 5500$ $NS_x^2 = 1085,\ NS_y^2 = 170,$ $NS_{xy} = 426,\ \Sigma Y^2 = 1450$	$N = 5,\ \Sigma X = 150,\ \Sigma Y = 80,$ $\Sigma XY = 2810,\ \Sigma X^2 = 5500$ $NS_x^2 = 1020,\ NS_y^2 = 175,$ $NS_{xy} = 400,\ \Sigma Y^2 = 1450$

TPA1 $b=\frac{426}{1085}=0.393^{*}$, $b_0=16-(0.393)(31)=3.82^{*}$ $E(Y,X)=(0.393)31/16=0.761^{*}$, $R^2=0.76^{*}$ all are incorrect estimates	**TPA1** $b=\frac{400}{1020}=0.392^{*}$, $b_0=17-(0.392)28=6.02^{*}$ (p.e 38%) $E(Y,X)=(0.392)28/17=0.64^{*}$, $R^2=32.5^{*}$ (p.e 18%) all are incorrect estimates
TPA2 $b=\frac{2810/5-31\cdot 16}{5500/5-31^2}=0.47^{*}$, $b_0=16-(0.47)(31)=1.43^{*}$ $E(Y,X)=(0.47)31/16=0.91^{*}$, $R^2=0.91^{*}$ all are incorrect estimates	**TPA2** $b=\frac{2810/5-28\cdot 17}{5500/5-28^2}=0.27^{*}$, $b_0=17-(0.27)28=9.44^{*}$ $E(Y,X)=(0.27)28/17=0.44^{*}$, $R^2=23.21^{*}$ all are incorrect estimates
OPA3 $b=\frac{25S_{xy}}{25S_x^2}=\frac{5\cdot 2810-150\cdot 80}{5\cdot 5500-150^2}=0.41$, $b_0=3.7, E(Y,X)=(0.41)150/80=0.768$, $R^2=0.988$ all are correct estimates	**OPA3** $b=\frac{2050}{5000}=0.41$, $b_0=3.7$ $E(Y,X)=0.768$, $R^2=0.988$ all are correct estimates

Case 5

Increasing the values of X by 2 we have:

$\bar{X}=32$ $\bar{Y}=16$ $e_1=2$ $e_2=1$ $\tilde{X}=34$ $\tilde{Y}=17$.

Note that $e_2 : e_1 = \bar{Y} : \bar{X}$; b_T $N=5$ $\Sigma X=160, \Sigma Y=80, \Sigma XY=2970,$

$\Sigma X^2=6120,\ \Sigma Y^2=1450,\ NS_x^2=1020,\ NS_y^2=175,\ NS_{xy}=420$

TPA1	**TPA2**
$b = \frac{420}{1020} = 0.41,$	$b = \frac{2970/5 - 34 \cdot 17}{6120/5 - 34^2} = 0.24^*,$
$b_0 = 17 - (0.41)34\ ;\ \ 3\,(\text{a.c.e.})$	$b_0 = 17 - (0.24)(34) = 8.84^*,$
$E(Y,X) = (0.41)34/17 = 0.82^*,$	$E(Y,X) = (0.24)34/17 = 0.48^*,$
$R^2 = 5.37^*$	$R^2 = 3.84^*$
only two are correct estimates	all are incorrect estimates

OPA3

$$b = 0.41 \qquad b_0 = 2.88$$

$$E(Y,X) = 0.82 \qquad R^2 = 0.99$$

all are correct estimates

** Incorrect estimates*
a.c.e.: almost correct estimate
p.e=percentage error

Table 2: *Baczkowski's Data Set*

Age	Price	Log(Price)	Age	Price	Log(Price)	Age	Price	Log(Price)
0	3972	3,5990	0	7596	3,8806	9	1695	3,2292
0	4600	3,6628	1	5350	3,7284	9	1995	3,2999
0	5541	3,7436	1	6475	3,8112	9	2250	3,3522
0	5591	3,7475	2	4450	3,6484	9	2250	3,3522
0	5817	3,7647	2	4995	3,6985	10	1395	3,1446
0	5909	3,7715	2	5495	3,7400	10	1425	3,1538
0	6641	3,8222	3	4995	3,6985	10	1595	3,2028
0	6723	3,8276	6	3250	3,5119	11	1295	3,1123
0	6951	3,8420	8	2495	3,3971	11	1695	3,2292

0	7246	3,8601	9	1495	3,1746	13	1295	3,1123

Table 3: *Computation of the regression coefficients in Regression of Lp on age*

$$\bar{X} = 4.5 \qquad \overline{Lp} = 3.537384$$

$$e_1 = 0 \qquad e_2 = -0.04$$

$$\tilde{X} = 4.5 \qquad \widetilde{Lp} = 3.5$$

$$N = 30,\ \Sigma X = 135,\ \Sigma Lp = 106.12,\ \Sigma X^2 = 1239,$$

$$\Sigma (Lp)^2 = 377.52,\ \Sigma X Lp = 441.95$$

$$\Sigma \left(X - \tilde{X} \right)^2 = 63.15,\ \Sigma \left(Lp - \widetilde{Lp} \right)^2 = 2.19,$$

$$\Sigma \left(X - \tilde{X} \right) \left(Lp - \widetilde{Lp} \right) = -35.58$$

TPA1

$$b = \frac{-35.58}{631.5} = -0.0563 \quad , \quad b_0 = 3.5 + (0.0563)(4.5) = 3.75^*$$

TPA2

$$b = \frac{444.95/30 - (4.5)(3.5)}{1239/30 - (4.5)^2} = 0.0483^* \ ,$$

$$b_0 = 3.5 + (0.0483)(4.5) = 3.71^*$$

OPA3

$$b = \frac{900\,S_{XLP}}{900\,S_x^2} = \frac{30(441.95) - 135(106.12)}{30 \cdot 1239 - 135^2} = -0.0563$$

$$30b_0 = 106.12 + (0.0563)(135) = 113.7265$$

$$\text{therefore} \qquad b_0 = 3.79$$

$$R^2 = \frac{30(3.79)(106.12) - (0.0563)30(441.95) - (106.12)^2}{30(377.52) - (106.12)^2} = 0.90$$

all correct estimates

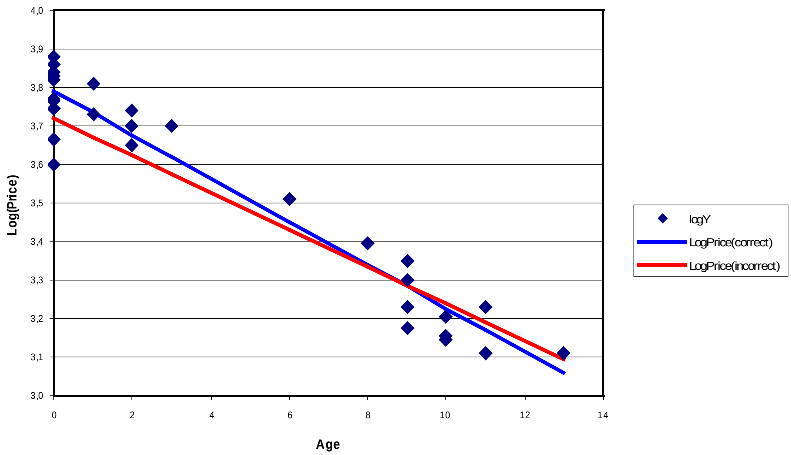

Figure 1. Scatterplot of Log(Price) given Age

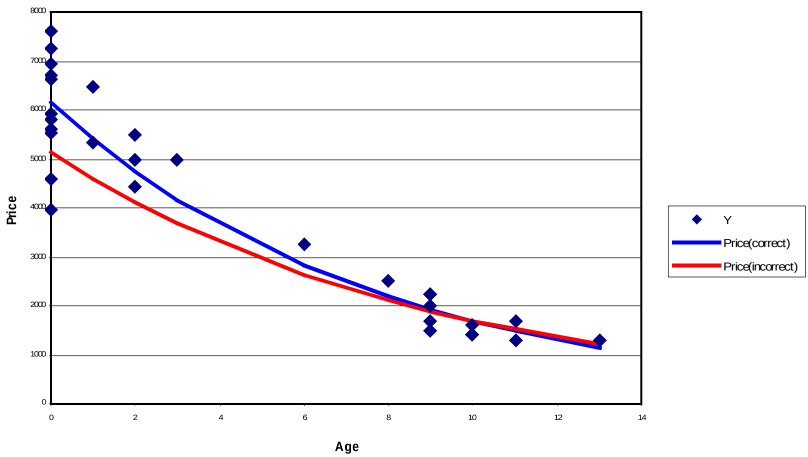

Figure 2. Scatterplot of Log(Price) given Age